4 Findings

4.1 Introduction

This chapter explores the answers this study provides to the four research questions put forward in 3.1. Each section deals with one of the questions. First, in 4.2, the free written response task data are explored to shed light on how second language learners start to use the two grammatical phenomena in focus. The analyses presented in 4.3 are intended to demonstrate how explicit and implicit instruction affect the learners' explicit and implicit knowledge of the target structures. Then, the focus shifts towards potentially interacting variables. In 4.4, the influence of developmental readiness, L1 similarity, and ID variables such as aptitude, motivation, learning style and age is examined. Finally, in 4.5, an exploration is provided of the potential influence of structure complexity.

4.2 Learning to use grammar

4.2.1 Introduction

RQ 1: How do second language learners develop the ability to use the target structures in spontaneous situations of second language use?

The analyses presented in this section are intended to describe how L2 learners develop the ability from 'not-knowing-how-to-use' to 'knowing-how-to-use' the degrees of comparison (DoC) and subordinate clauses (SubC) in contexts of free use. This will be investigated irrespective of potential effects of instruction. There are two main avenues of exploration to do with the development of more efficient use of the structure and rate of learning. In 4.2.2, efficiency of use will be examined by means of the emergence of correct use of the target structures in relation to how they are realized in the utterance. To this end, descriptive statistics are presented that provide a static description of the use of the target structures at each time of measurement (T0, T1, and T2), differentiated

according to their realizations. The realizations that will be distinguished have already been introduced in 3.6.4. Subsection 4.2.3 focuses on how fast the ability to use the target structure develops. The analyses will be presented for each target structure separately, and all participants are included.

4.2.2 Structure realizations and correct use

Degrees of comparison

Before correct use of the DoC will be considered in relation to their realizations, it is helpful to discuss some general characteristics of use. Table 4.1 describes the use of the DoC at T0, T1, and T2 without further differentiation. In total, the free written response task provided 18 contexts that invited the participants to use the DoC. The table shows that overall use increases from an average of 6.4 instances of use in the pre-test to 8.2 in the second post-test. Decomposing this figure into correct and incorrect use shows that the DoC were used correctly more often than incorrectly. Correct use increases from 4.7 times at T0 to 7.0 times at T2. Incorrect use occurred on average just over one time per test. Sometimes, the participants avoided the use of the DoC by means of circumscription (for example by using *meer dik* (more fat) instead of *dikker* (fatter); see 3.6.4). Avoidance did not occur very often: it was used on average 0.7 times at T0, and its use dropped to 0.2 times at T2. Finally, productivity expresses the number of different adjectives with which a form of the DoC was used. The average was 3.3 at T0, and it increased to 4.7 at T2.

The question explored here is whether correct use depends on the structure's realization. In 3.6.4, the following realizations have been put forward as relevant to the use of the DoC: 1) the degree of comparison expressed (comparative or superlative); 2) the syntactic form of the comparison: attributive, predicative, or predicative including a comparative clause; 3) and the adjective used. Correct use of the DoC will be considered in relation to each these realizations.

First, differences in correct use of comparative and superlative realizations will be considered. The free written response task intended to elicit both forms, without expressly intending to elicit the use of either comparative or superlative forms in particular situations. Table 4.2 provides mean use figures for each point of measurement. In addition, percentages are presented that express how often the particular form was used correctly. The numbers clearly show that the subjects had a clear preference for using the comparative form in this test. At

TABLE 4.1 *The use of the degrees of comparison in the free written response task.*

Characteristics of use	***Statistic***	***Use at T0 (N=101)***	***Use at T1 (N=101)***	***Use at T2 (N=76)***
Total use	*Mean*	6.4	8.1	8.2
	SD	4.2	4.5	4.1
	Range	0 – 21	0 – 22	1 – 19
Correct use	*Mean*	4.7	6.6	7.0
	SD	4.3	4.9	4.4
	Range	0 – 18	0 – 22	0 – 17
Incorrect use	*Mean*	1.7	1.5	1.2
	SD	1.9	2.0	1.7
	Range	0 – 8	0 – 9	0 – 8
Avoided use	*Mean*	0.7	0.4	0.1
	SD	1.1	0.9	0.4
	Range	0 – 5	0 – 4	0 – 2
Productivity	*Mean*	3.3	4.3	4.7
	SD	2.8	2.8	2.7
	Range	0 – 12	0 – 10	0 – 10

TABLE 4.2 *Mean use and correct use in percentages of comparative and superlative Doc in the free written response task.*

	Use at T0		***Use at T1***		***Use at T2***	
Realization	***M (SD)***	***% cor***	***M (SD)***	***% cor***	***M (SD)***	***% cor***
Comparative	5.4 (3.9)	71	6.9 (4.5)	80	7.0 (3.9)	84
Superlative	1.0 (1.3)	77	1.3 (1.6)	83	1.1 (1.3)	90

T0, it was used 5.4 times, while the superlative was used on average only once. There was a clear rise in mean use of comparative forms from pre-test to second post-test; however, mean use of superlative forms remained more or less constant. This may indicate either that there was no learning effect for superlatives, or that the task simply did not invite the use of superlatives sufficiently. If one considers the percentages of correct use, there do not seem to be any differences between comparatives and superlatives. Both were used correctly in more than 70 percent of the cases at T0, and both demonstrate an increase in correct use.

Another realization that may have led to differentiation in correct use is the syntactic form in which instances of the DoC appear. In 3.6.4, these realizations have already been introduced: for comparative DoCs, predicative use (Pred) and predicative use including a comparative clause (Pred+) have been distinguished. For superlatives, attributive (Att) and predicative (Pred) use have been distinguished. Again, mean use figures at each time of measurement and correct use in percentages are presented (Table 4.3). Both Pred and Pred+ types of comparatives occurred regularly at T0: Pred use occurred 2.9 times, and Pred+ occurred 2.6 times. For both, mean use increased from T0 to T2. Percentages of correct use also increased substantially for both. However, there is a clear difference in correct use of Pred and Pred+. At T0, the former type was used correctly in 83 percent of the instances of occurrence, while 50 percent of the latter type was used correctly: a difference of 33 percentage points. At T2, this difference was reduced to 16 percentage points.

Further differentiation of superlative instances of use led to very low frequencies of use. In addition, there was no noticeable increase in use in either attributive or predicative realizations. If percentages of correct use are considered, however, there does seem to be increased performance. In fact, at T2, the L2 learners made very few mistakes when using a superlative form. There are no indications of differentiated correct use between attributive and predicative superlatives.

The final realization to be considered is the adjective itself. In total, the free written response task elicited the use of the DoC with 29 different adjectives. Some of these occurred rather frequently, while others occurred only once or twice. Obviously, the situations used to elicit the DoC favour the use of particular adjectives, while others will not tend to occur as much. Therefore, frequencies of use are not informative. Table 4.4 presents percentages of correct use of the most frequently occurring adjectives (those that occur at least 25 times at each point of measurement) in the free written response task data.

Again, the question is whether one can speak of differentiated correct use depending on the adjective used. What stands out from Table 4.4 is that there did not seem to be major differences in correct use at each time of measurement. *Groot* and *oud* were the most difficult to use correctly, judging from their scores at T0. The difference between the highest and lowest score at T0 was 22 percentage points (the difference between *dik* and *groot*). This difference levelled out somewhat at T1 and T2; and at T2, all scores fell within a margin of 14 percentage points. For all adjectives, percentages of correct use increased from T0 to T2, although there was one exception to the rule: *dik* showed a slight decrease.

TABLE 4.3 *Mean use and correct use in percentages of the DoC in the free written response task, differentiated according to syntactic realizations.*

	T0 (N=101)		T1 (N=101)		T2 (N=76)	
Realization	***M (SD)***	***% cor***	***M (SD)***	***% cor***	***M (SD)***	***% cor***
Comparative [a]						
Pred.	2.9 (2.7)	83	3.8 (2.8)	89	3.6 (2.7)	92
Pred+CompC	2.6 (2.0)	50	3.2 (2.6)	68	3.3 (2.5)	76
Superlative						
Att.	0.3 (0.6)	81	0.4 (0.7)	86	0.2 (0.4)	94
Pred.	0.8 (1.5)	83	1.0 (1.2)	82	0.9 (1.1)	87

a – In principle, attributive comparatives can also occur in Dutch, but they hardly did in these data. Therefore, they were not included in this table.

TABLE 4.4 *Correct use of the degrees of comparison in percentages for eight frequently occurring adjectives.*

	Correct use in %		
Realization	***T0 (N=101)***	***T1 (N=101)***	***T2 (N=76)***
dik (fat)	84	76	79
duur (expensive)	69	67	78
goedkoop (cheap)	76	86	88
groot (big)	62	76	84
klein (small)	76	81	89
mooi (beautiful)	76	77	92
oud (old)	64	75	80
snel (fast)	82	81	91

All in all, the analyses presented here do not provide many indications of differentiated correct use across different realizations of the DoC. The clearest indication of differentiated use was observed between Pred and Pred+ comparatives. There may have been differentiated use between some of the adjectives, too.

Subordinate Clauses

Table 4.5 displays the general characteristics of use of the subordinate clauses. It provides mean use, standard deviation and the range at each time of

TABLE 4.5 *The use of subordinate clauses in the free written response task.*

Characteristics of use	*Statistic*	*Use at T0 (N=101)*	*Use at T1 (N=101)*	*Use at T2 (N=76)*
Total	*Mean*	13.6	13.5	14.1
	SD	5.9	5.8	5.9
	Range	1 – 25	0 – 22	0 – 24
Correct	*Mean*	5.0	6.4	8.2
	SD	5.2	5.4	5.5
	Range	0 - 20	0 – 18	0 – 20
Incorrect	*Mean*	7.6	6.0	4.8
	SD	5.4	4.9	4.7
	Range	0 – 18	0 – 18	0 – 17
Avoided	*Mean*	2.7	2.8	2.7
	SD	4.4	4.6	4.6
	Range	0 – 15	0 – 16	0 – 15

TABLE 4.6 *Mean use and correct use in percentages of conditional and causal SubCs in the free written response task.*

	T0 (N=101)		*T1 (N=101)*		*T2 (N=76)*	
Realization	*M (SD)*	*% cor*	*M (SD)*	*% cor*	*M (SD)*	*% cor*
Conditionals	3.0 (1.4)	64	3.0 (1.2)	72	3.2 (1.2)	85
Causals	9.6 (5.3)	28	9.4 (5.3)	40	9.7 (5.5)	49

measurement, and broken down according to total, correct, incorrect, and avoided use. In total, the free written response task provided 21 opportunities to use SubCs. On average, the participants actually took this opportunity 13.6 times at T0. As can be seen, total mean use of SubCs remained more or less constant from T0 to T2. However, if one considers correct use, there was an increase in mean use from 5.0 at T0 to 8.2 at T2. Conversely, incorrect use decreased from 7.6 at T0 to 4.8 at T2. As the total use figure is made up of correct and incorrect use, the lack of increase in total use must be due to these opposite trends in correct and incorrect use. In 3.6.4, it was pointed out that Dutch causal subordinate clauses can easily be avoided by means of a coordinate construction with *want*, which does not affect word order. The average use of avoidance was 2.7 times at T0. There was no noticeable increase or decrease in the use of this construction.

Again, correct use of subordinate clauses will be examined in relation to how the structure was realized in the utterance. The realizations to be considered have been identified in 3.6.4: for SubC, the contexts of use are defined by: 1) the type of subordinate clause (conditional or causal); 2) verb phrase complexity; and 3) clause complexity. Each will be discussed below.

The first type of realization to be considered is the subordinate relation the clause expresses. It is important to remember that the test primarily invited the use of conditional and causal relationships, and only these two types are contrasted here (see 3.3.3). In addition, it should be kept in mind that the test contained only three situations eliciting the use of conditionals, while it contained 18 situations targeting causal SubCs because of expected difficulties in eliciting their use. Table 4.6 shows that average use of conditional SubCs was approximately three times per test at each time of measurement. Causal SubCs were used somewhere between 9.4 and 9.7 times. There was no increase in mean use. If one considers the percentages of correct use, a rather large difference in correct use between conditionals and causals shows up. At T0, 64 percent of the conditionals were used correctly, a number that increased to 85 percent at T2. In contrast, no more than 28 percent of the causal SubCs were used correctly at T0; and despite an increase in correct use, this difference still amounted to 36 percentage points at T2.

Another realization to be considered pertains to the complexity of the verb phrase used in SubCs. The verb phrases distinguished were those consisting of 1) single auxiliaries (Aux); 2) single verbs (SV); and multiple verbs (MV) (see also 3.6.4). Table 4.7 displays the frequencies of use and percentages of correct use from T0 to T2 for each type of verb phrase, and – given the previous analyses – differentiated according to type of SubC. The differences in mean use were substantial, but – as pointed out before – one should not attach too much value to these, as these may well be influenced by the nature of the test. The percentages show that the students obtained progress, irrespective of verb phrase complexity. However, for both conditional and causal SubCs, auxiliaries were used correctly most often, although the difference with single-verb phrases was not very large. For both conditional and causal SubCs, the students clearly had most difficulties with verb phrases containing multiple verbs.

Finally, the complexity of the subordinate clause may also affect the L2 learner's ability to properly place the verb. Three different types of clause structure were coded: those consisting of adjectives (Adj); direct objects (Do); and complex constituent structures (CompC) (also see 3.4). Again, the statistics will

TABLE 4.7 *Mean use and correct use in percentages of conditional and causal SubCs in the free written response task, differentiated for verb phrase complexity.*

	T0 (N=101)		T1 (N=101)		T2 (N=76)	
Realization	***M (SD)***	***% cor***	***M (SD)***	***% cor***	***M (SD)***	***% cor***
Conditionals						
Aux	0.8 (0.8)	74	0.7 (0.7)	90	0.9 (0.8)	90
SV	1.7 (1.3)	79	1.6 (1.1)	83	1.7 (1.1)	90
MV	0.5 (0.7)	50	0.6 (0.7)	68	0.4 (0.5)	68
Causals						
Aux	5.4 (2.9)	42	5.0 (2.6)	56	4.8 (3.0)	63
SV	2.5 (1.8)	34	2.4 (1.6)	43	2.5 (2.2)	50
MV	0.9 (0.8)	15	1.7 (1.5)	27	1.5 (1.8)	38

TABLE 4.8 *Mean use and correct use in percentages of conditional and causal SubCs in the free written response task, differentiated for clause complexity.*

	T0 (N=101)		T1 (N=101)		T2 (N=76)	
Realization	***M (SD)***	***% cor***	***M (SD)***	***% cor***	***M (SD)***	***% cor***
Conditionals						
Adj	0.9 (0.7)	74	1.0 (0.8)	85	1.2 (1.0)	93
Do	1.6 (1.1)	74	1.3 (0.9)	83	1.4 (1.1)	84
CompC.	0.5 (0.5)	55	0.5 (0.5)	73	0.5 (0.4)	76
Causals						
Adj	5.1 (2.8)	45	5.2 (2.9)	54	5.0 (2.7)	59
Do	3.1 (2.0)	28	2.8 (1.9)	41	2.5 (1.8)	53
CompC.	1.0 (1.0)	16	1.2 (1.0)	27	1.3 (1.2)	37

be presented for conditional and causal SubCs separately (see Table 4.8). There will be no further differentiation according to the kinds of verb phrases used, because frequencies of occurrence would be too low for such a fine-grained differentiation. The mean use figures serve to provide an indication of frequency of use of the different types of SubCs. What stands out is that mean use remained more or less constant throughout the time. However, the percentages of correct use again rose without exception. For both conditional and causal SubCs, Adj clauses were easiest to realize correctly. In fact, at T2, 93 percent of these were

TABLE 4.9 *Co-occurrence of conditional and causal subordinate clauses.*

Characteristics of use	*T0 % of N (N = 101)*	*T1 % of N (N = 101)*	*T2 % of N (N = 76)*
No correct conditionals, no correct causals	16	11	3
Correct conditionals, no correct causal	30	22	15
No correct conditionals, correct causals	3	3	4
Correct conditionals and causals	51	64	78

used correctly in case of conditional SubCs. The differences in correct use between Adj clauses and Do clauses were marginal. As to be expected, placing the verbs correctly clearly proved most difficult in CompC clauses.

For SubCs, one additional analysis was performed. Given the clear differences that were found in correct use between conditional and causal SubCs, it is interesting to compare and contrast simultaneous use of each type. Table 4.9 describes the appearance of correct use of causals and conditionals. At T0, 16 percent of the subjects did not use any correct conditionals and causals. As would be expected, this number dropped to three percent at T2. The table also shows that correct use of verb placement in conditional clauses is by no means a guarantee that the L2 learner will also place the verb correctly in clauses expressing causal relationships. At T0, 30 percent used correct instances of conditionals, but no correct causals. This number decreased to 15 percent at T2. However, correct verb placement in causals almost always entailed correct verb placement in conditionals. Only 3 to 4 percent of the subjects used causals correctly without using conditionals correctly. The ability to use verb placement correctly in both contexts is demonstrated by 51 percent of the subjects, and increasing to 78 percent at T2.

In sum, these analyses provide a number of indications that correct use of subordination depends on its realization. There was a substantial difference in correct use of conditional and causal SubCs, and the ability to use conditionals correctly clearly preceded the ability to causals correctly. The complexity of the verb phrase and the sentence structure also led to differentiated use.

4.2.3 Developing the ability to use the target structure

This subsection focuses on how fast L2 learners move from not-knowing-to-use a particular grammar structure to knowing-to-use it. This will be explored by investigating the use of the target structures with reference to four stages of

development, which are based on the developmental stages White (1998) used to monitor progress in the use of possessive determiners. On the basis of previous research, White was able to define a rather fine-grained sequence of eight stages, pertaining specifically to the development of possessive determiners. These eight stages roughly reflect four overarching stages: preemergence, emergence, postemergence and targetlike performance (White, 1998: p. 105). Preemergence simply refers to non-use or avoidance of the target structure. In the emergence stage, learners start using the target structure, but incorrectly. White interprets this as a phase in which the language learner does not yet apply the rule. In the postemergence stage, learners slowly start to use the rule correctly; and in the last stage, the rule is used in a targetlike way. These stages, then, are not structure-specific, and they are concrete and observable in data. The ability to use a target structure in spontaneous situations moves from: a stage of 1) no use; to 2) incorrect use; to 3) variable use; and 4) all correct use. The expectation that the ability to use the target structures correctly develops slowly will be addressed by means of these stages. There will be no further differentiation according to contexts of use for the DoC. For the subordinate clauses, the analyses will be performed separately for conditional and causal SubCs.

Degrees of comparison

If the participants of this study are categorized according to how they use the target structure at each time of measurement, it is possible to see how they move from one stage to the next. The top half of Table 4.10 depicts how the learners use the DoC at T0 and T1. At T0, for instance, seven students did not demonstrate any use of the DoC, but at T1, five of them actually started to use the DoC in the free written response task: three show variable use, and two use all forms of the DoC correctly. Out of the eleven students that used the DoC only incorrectly at T0, six demonstrated variable use at T1. Most students showed variable use at T0: 50 out of 101. At T1, 20 started to use the DoC without error, and only two students 'moved backwards', in that they did not use any correct forms anymore at T1, while they did at T0. The same kind of 'deterioration' can be observed for the 33 students that did not make any errors at T0. At T1, 12 showed signs of variable use. If one considers moving from one stage to the next an indication of progress, then 31 students can be said to demonstrate increased performance. Fourteen students showed a decrease in performance: this number is almost entirely caused by students that demonstrated variable use at T1 while

TABLE 4.10 *Changes in patterns of use of the degrees of comparison in the free written response task.*

		Patterns of use at T1 in absolute numbers (N = 101)			
Patterns of use	***Use at T0***	***No use***	***incorrect use***	***Variable use***	***Correct use***
No use	7	2	-	3	2
Only incorrect use	11	-	5	6	-
Variable use	50	-	2	28	20
Only correct use	33	-	-	12	21
	Use at T1	***Patterns of use at T2 in absolute numbers (N = 76)***			
No use	1	-	-	1	-
Only incorrect use	2	-	2	-	-
Variable use	39	-	1	25	13
Only correct use	34	-	-	10	24

making no errors at T0. A large majority, 56 students out of 101, remained in the same stage of development. Although cell frequencies are lower due to mortality in the research sample, the lower half of Table 4.10 yields the same picture. There was some progression, primarily from variable use to correct use; there was also regression from correct use to variable use; but the majority again did not change in how they used the degrees of comparison.

Subordinate clauses

The same stages of development can be applied to students' use of subordinate clauses. Tables 4.11 and 4.12 present the changes in patterns of use for respectively conditional and causal SubCs. For both types of SubCs, the picture that emerges is very similar to the one described for the DoC. The number of students that showed progression, either from T0 to T1 or from T1 to T2, was always substantially larger than the number of students that showed regression. But the majority of the students – invariably more than 50 percent – did not show any difference in the way they used subordinate clauses. Finally, the category that again was least stable was the 'correct use' category: although no errors were made at T0, a substantial amount of students returned to variable use at T1.

TABLE 4.11 *Changes in patterns of use of conditional subordinate clauses in the free written response task.*

		Patterns of use at T1 in absolute numbers (N = 101)			
Patterns of use	*Use at T0*	*No use*	*incorrect use*	*Variable use*	*Correct use*
No use	5	3	2	1	-
Only incorrect use	41	1	4	6	3
Variable use	45	-	4	15	16
Only correct use	10	-	-	11	35
	Use at T1	*Patterns of use at T2 in absolute numbers (N = 76)*			
No use	4	2	-	-	-
Only incorrect use	27	1	1	4	2
Variable use	37	-	2	6	15
Only correct use	8	-	-	6	38

TABLE 4.12 *Changes in patterns of use of causal subordinate clauses in the free written response task.*

		Patterns of use at T1 in absolute numbers (N = 101)			
Patterns of use	*Use at T0*	*No use*	*incorrect use*	*Variable use*	*Correct use*
No use	6	1	2	-	2
Only incorrect use	14	2	22	16	1
Variable use	35	1	3	35	6
Only correct use	46	2	-	4	4
	Use at T1	*Patterns of use at T2 in absolute numbers (N = 76)*			
No use	2	3	-	-	1
Only incorrect use	7	1	5	12	-
Variable use	23	1	1	36	6
Only correct use	44	3	-	2	5

4.2.4 Conclusion

Does correct use depend on the realization of the target structure? The question wase addressed for each target structure separately. Characteristic of the use of the DoC was that the frequency of use in the free written response task increased from T0 to T2. When total use scores were broken down according to correct and incorrect use, then incorrect use of the DoC turned out to quite infrequent. Correct use was more frequent, and a clear increase in correct use was observed. Productivity, the number of different adjectives the DoC were used with, also showed a clear increase from T0 to T2. Correct use of the DoC was investigated in three different contexts. No differences have been observed in correct use of comparatives or superlatives, nor have clear differences been found based on the adjectives with which the DoC were used. However, correct use of comparatives did seem to be affected by the presence of comparative clauses. With respect to this latter finding, it is important to point out that omission of the comparative suffix in the presence of a comparative clause can be clearly marked as incorrect. In contrast, omission of this suffix in all other contexts would go unnoticed, because the grammaticality of the sentence is not affected (see also 3.6.4). For this reason, the difference in the percentages of correct use of comparatives with and without a comparative clause may be the result of unnoticed errors coded as *no use*. The conclusion must therefore be that it has not been demonstrated beyond doubt that there is differentiation in correct use according to the structure's realization.

The mean use of SubCs remained constant from T0 to T2. However, there was a clear increase in mean correct use and a decrease in mean incorrect use. For SubCs, the structure's realization does seem to make a difference. Especially the difference in percentages of correct use between conditional and causal SubCs is remarkable: conditionals were used correctly much more. Further differentiation according to verb phrase complexity also revealed differences in correct use between phrases consisting of auxiliaries or single verbs on the one hand and phrases consisting of multiple verbs on the other. Auxiliaries were easiest to use; phrases containing multiple verbs were clearly more difficult to use correctly. Similarly, when differentiating for clause complexity, a difference in correct use was observed between relatively simple and more complex clauses. Percentages of correct use were lower as clause complexity increased. The most pervasive difference was observed between conditional and causal SubCs, though. For this reason, their co-occurrence was investigated as well. This clearly indicated that L2 learners started using conditionals correctly before causals, and correct use of

causal SubCs seemed to preclude the possibility that conditionals were used incorrectly.

In order to gain insight into how fast the participants of this study developed the ability to use the target structure, students were categorized according to four developmental stages: preemergence, emergence, postemergence and targetlike behaviour. The analyses have shown that the participants of this study did not show much progress in terms of these stages. The participants were found to be very persistent in how they used the target structures, and most did not change their behaviour from one point of time to the next. As the stages used were not structure-specific, they may simply have been too crude to capture progress. Nevertheless, as this study spanned a period of three to four months, it seems justified to conclude that the ability to use the two grammatical phenomena under study in spontaneous situations develops slowly.

4.3 FFI, explicit knowledge, and grammatical development

4.3.1 Introduction

The analyses presented in this section pertain to the second research question:

RQ 2: How are explicit and implicit FFI related to the development of explicit and implicit grammatical knowledge?

The expectation was that explicit instruction is superior only in promoting the development of explicit grammatical knowledge. For the development of implicit grammatical knowledge, no differences were expected for the two types of instruction. These hypotheses were tested by means of univariate repeated measures ANOVA's for two different measures of grammatical progress. The untimed grammaticality judgement (GJ) task scores were used as an indication of explicit progress, while correct use of the target structures in the free written response task was taken to be an indicator of implicit progress. An interaction was sought between instruction (Instr), operationalized as a three-level between subjects factor (EI, II and NI); and progress, also a three-level within subjects factor (T0, T1, and T2). The analyses were performed separately for each target structure, and each measure of progress.

4.3.2 FFI and the development of the degrees of comparison

Preliminary analyses

For each target structure at each time of measurement, means and standard deviations have been computed. Table 4.13 depicts these. As pointed out in 3.4.3, for the degrees of comparison, there were insufficient control group data because a number of control group participants received explicit instruction in the degrees of comparison. For this reason, the analyses for the degrees of comparison were limited to a comparison of the EI and II conditions. The pre-test data of both measures were submitted to a one-way analysis of variance to test for significant differences between the groups at the start of the experiment. No significant differences were found. In addition, the three groups were tested for significant differences on any of the general proficiency measures (see 3.4) and ID variables (see 3.5). No differences were found, which is an indication that the participants were randomly distributed across the conditions.

As no significant differences are expected between instruction and correct use in the free written response, sufficient statistical power to detect interaction effects is an important design requirement. Assuming equal group sizes of 32 per group for the sake of power analysis, the 2 by 3 repeated measures design tested in this subsection reaches a power of 1.0 to detect medium within-subjects effects and medium interaction effects (f = .25, according to Cohen's (1988) standards).

Main analyses

For each measure, a 2 by 3 ANOVA was run, comparing only the EI and II groups. The results of these analyses can be found in Table 4.14. The table shows that progress over time was significant for both the explicit and the implicit

TABLE 4.13 *Degrees of comparison: means and standard deviations for GJ task scores and correct use in the free written response task.*

			T0	*T1*	*T2*
Measure	*Instr*	*n*	*M (SD)*	*M (SD)*	*M (SD)*
GJ task	EI	32	2.44 (2.08)	3.50 (1.93)	3.78 (1.98)
	II	35	1.94 (1.78)	2.06 (2.18)	2.49 (2.02)
Correct use	EI	32	5.44 (4.79)	7.53 (5.91)	7.25 (4.79)
	II	35	5.09 (4.72)	6.66 (4.54)	7.14 (4.37)

TABLE 4.14 *Degrees of comparison; Repeated measures ANOVA for progress and instruction*

Source	*SS*	*df*	*MS*	*F*	*p.*
GJ task [a]					
progress	30.36	1.65	18.45	13.38	.00
progress x Instr	8.71	1.65	5.29	3.39	.03
Error	147.45	106.96	1.38		
Correct Use					
progress	158.52	2	79.26	12.84	.00
progress x Intr	5.13	2	2.57	.42	.66
Error	802.42	130	6.17		

a – Because the assumption of sphericity was violated, the Greenhouse-Geisser estimate is reported for this measure.

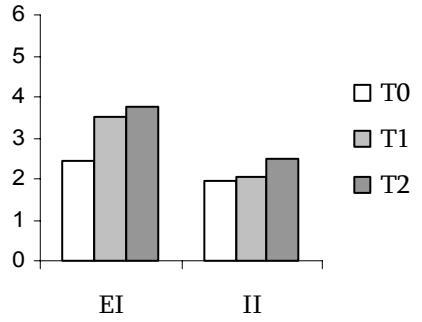
knowledge measure. However, for Correct Use, there was no interaction between progress and the type of instruction received. Only the grammaticality judgement (GJ) task scores were affected by the instruction provided: explicit instruction led to higher scores on the GJ task. These results will be discussed in more detail based on Figure 4.1, which illustrates per measure how the groups scored at each time of measurement. In addition, the main and interaction effects found will be explored further by means of post hoc analyses.

As pointed out in 3.4.3, this study suffered from considerable mortality in the subject sample from T1 to T2. To assess whether this loss of subjects may have affected the outcome, a 2 by 2 ANOVA has been run as well, in which the within subjects factor progress only had two levels, T0 and T1. The results were the same: significant overall progress was observed for both the GJ task and Correct Use, while an interaction between progress and instruction was found only for the GJ task scores. Thus, mortality does not seem to have affected these results.

Figure 4.1A indicates that the main effect for progress on the GJ task is mainly due to the progress obtained by the EI group: the II group showed marginal progress only, which would probably not differ much from a NI group. Post hoc analyses investigating progress for each group individually support this observation. Paired samples t-tests revealed that the only significant progress was obtained by the EI group between T0 and T1($t(31) = -3.14$, $p < .01$). Post hoc analyses have also been used to investigate the interaction between progress and instruction. One-way ANOVA's revealed significant differences between the two

FIGURE 4.1 *Mean scores for the grammaticality judgement task and correct use in the free written response task at each time of measurement.*

A: GJ task mean scores:



B: Correct use mean scores:

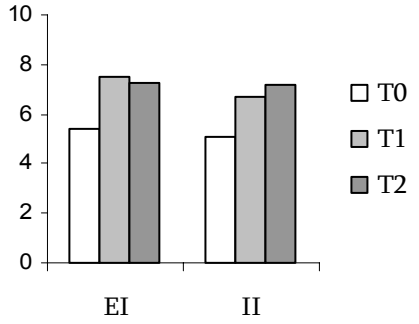


TABLE 4.15 *Explicit DoC knowledge: post hoc pairwise comparisons at each time of measurement.*

Time of measurement	***Observed difference*** [a]	***F***	***Effect size (in d)***
T0	EI = II	-	-
T1	EI > II	$F(1,65) = 8.14$	.70
T2	EI > II	$F(1,65) = 7.01$	.64

a – significant differences at alpha = .05 and using the Bonferroni adjustment are indicated by > or <.

groups at both T1 and T2, the EI significantly outperforming the II group at both times. In Table 4.15, these analyses are reported along with the effect sizes in case of significant differences.

Figure 4.1B shows that progress in *correct use* was similar for the EI and II groups. It also suggests that progress was obtained primarily between T0 and T1. From T1 to T2, which actually spans a longer time period, the students obtained only marginal progress. This observation was tested and confirmed by means of paired samples t-tests. Progress was significant from T0 to T1 (*correct use*: $t(66) = -3.98$), but no significant progress was obtained between T1 and T2. Overall, the size of the effect from T0 to T2 was moderate ($d = .41$) in terms of Cohen's (1988) standards.

4.3.3 FFI and the development of subordinate clauses

Preliminary analyses

In this subsection, the same analyses are presented as in the previous subsection, but this time measures of SubC progress are used. Table 4.16 summarizes the means and standard deviations for each measure at each time of measurement. One-way ANOVAs were used to assess for both measures whether the groups differed significantly at T0. No significant differences were observed. In addition, the three groups were tested for significant differences on any of the general proficiency measures (see 3.4) and ID variables (see 3.5). Again, no differences were found. Although the NI group is small, the current 3 by 3 repeated measures design still reaches sufficient power of to detect medium within-subjects and interaction effects: if all groups consisted of nine participants, the power to detect an interaction effect size of f = .25 would still be .98.

Main analyses

For each measure, 3 by 3 ANOVA was run in order to compare the progress obtained by the three different instruction groups. The results are displayed in Table 4.17. The picture is similar to that of the DoC: main effects were observed for both the explicit and implicit measures, but an interaction between progress and instruction was found only for the GJ task. Thus, the instruction does not

TABLE 4.16 *Subordinate clauses: means and standard deviations for GJ task scores and correct use in the free written response task.*

			T0	*T1*	*T2*
Measure	*Instr*	*N*	*M (SD)*	*M (SD)*	*M (SD)*
GJ task	EI	35	2.17 (2.41)	3.86 (2.64)	3.63 (2.69)
	II	32	1.44 (1.72)	2.44 (2.26)	3.19 (2.47)
	NI	9	1.00 (0.71)	0.56 (0.88)	0.67 (0.87)
Correct use	EI	35	5.94 (5.75)	7.03 (5.52)	7.54 (4.90)
	II	32	6.09 (5.42)	8.19 (5.39)	9.25 (5.94)
	NI	9	4.11 (4.96)	3.56 (4.28)	6.00 (6.12)

TABLE 4.17 *Subordinate clauses; Repeated measures ANOVA for progress and instruction*

Source	*SS*	*df*	*MS*	*F*	*p.*
GJ task					
progress	26.68	2	13.34	6.60	.00
progress x Instr	26.93	4	6.73	3.33	.01
Error	295.17	146	2.02		
Correct use [a]					
progress	131.07	1.75	75.04	9.33	.00
progress x Instr	40.66	3.49	11.64	1.45	.23
Error	1026.01	127.51	8.05		

a – Because the assumption of sphericity was violated, the Greenhouse-Geisser estimate is reported for this measure.

seem to affect performance on the free written response task. Only the GJ task scores are affected by the instruction. Mortality was again not found to have affected the results presented here. Using a two-level factor for progress, thus increasing the subject sample to 101, the same outcome was obtained: the loss of subjects between T1 and T2 is not likely to have affected the results substantially.

The analyses presented in section 4.2.2 have shown that the participants of this study made a differentiation in correct use of conditional and causal SubCs. In order to test the possibility that the instruction affected the use of either conditional or causal SubC development separately, the analyses have also been conducted for correct use of each type of SubC. The results yielded the same picture: for both types of SubCs, main effects were found, but no interaction between correct use and instruction. The results presented in Table 4.17 will be further explored by means of figure 4.2 and post hoc analyses, without making the differentiation between conditional and causal SubCs.

For the GJ task, a main effect and an interaction between progress and instruction were observed. Figure 4.2A clearly suggests that the overall main effect for time was caused by both experimental groups. The control group simply did not show progress on grammaticality judgement performance. Again, paired samples t-tests have been used to examine progress for each group individually. Significant progress was obtained by the experimental groups progress between T0 and T1 (for the EI group: $t(34) = -4.89$, $p < .01$, and for the II group ($t(31) = -3.09$, $p < .01$). This is an indication that little additional progress was obtained after the subjects stopped receiving instruction.

FIGURE 4.2 *Mean scores for the grammaticality judgement task and correct use in the free written response task at each time of measurement.*

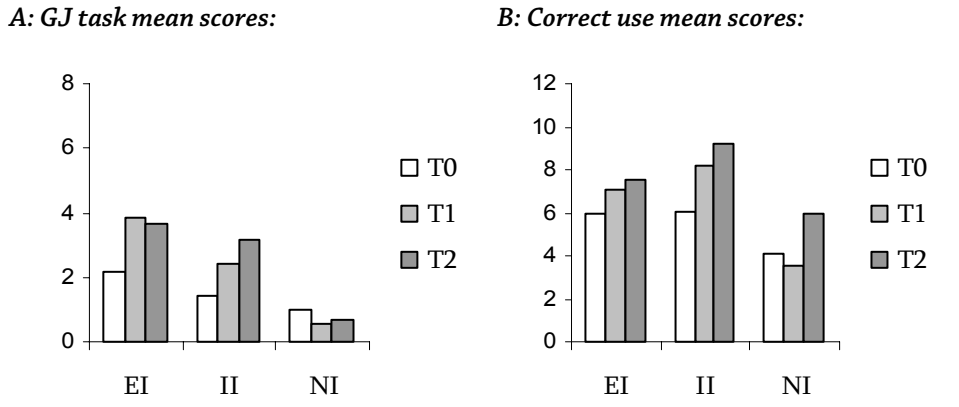


TABLE 4.18 *Explicit SubC knowledge: post hoc pairwise comparisons at each time of measurement.*

Time of measurement	*Observed difference* [a]	*F*	*Effect size (in d)*
T0	EI = II = NI	-	-
T1	EI = II	-	-
	EI > NI	F (1,42) = 13.52	1.37
	II = NI	-	-
T2	EI = II	-	-
	EI > NI	F (1,42) = 10.46	1.21
	II > NI	F (1,39) = 8.93	1.13

a – significant differences at alpha = .05 and using the Bonferroni adjustment are indicated by > or <.

The interaction found between progress and instruction for the GJ task indicates differentiated progress across the conditions. Figure 4.2A also clearly demonstrates that this interaction was to a large extent caused by the control group behaving differently. Indeed, post hoc pairwise comparisons using the bonferroni adjustment reveal a number of significant differences. Table 4.18 summarizes the outcomes. Before the instruction, no significant differences were

observed between the groups. Immediately after the instruction, the EI group significantly outperformed the NI group. At the third time of measurement, the II group significantly outperformed the NI group too.

Main effects were observed for *Correct Use*. These effects were further explored by means of paired samples t-tests. From T0 to T1, the progress that was overall obtained was significant (correct use: $t(75) = -3.29$, $p < .01$). This also applied to progress between T1 and T2: (correct use: $t(75) = -2.55$, $p < .01$). Thus, subjects increased their implicit knowledge during the instruction, and continued to do so in the two months following. The size of the effect from T0 to T2 was again moderate ($d = .42$).

4.3.4 The interface between explicit and implicit knowledge

Introduction

The analyses presented in the previous subsections (4.3.2 and 4.3.3) have clearly demonstrated for both target structures that received explicit instruction did not present an advantage to students for developing implicit grammatical knowledge. This was notwithstanding the fact that the groups differed with respect to their explicit knowledge of the target structures. Thus, this study does not lend support to an interface between explicit and implicit knowledge; and similarly, it does not provide arguments in favour of explicit FFI. However, it may be that L2 learners differ in their ability to handle and benefit from explicit instruction, and that this is only given to some, not all L2 learners. In other words, it is possible that the presence of an interface between explicit and implicit knowledge is obscured in these data by those L2 learners that did not benefit from the explicit instruction. To investigate this, the analyses were rerun, comparing an 'explicit knowledge' (EK) group with a 'no explicit knowledge' (NEK) group.

The EK and NEK groups were created taking the EI and the II groups as a starting point. From the EI group, all participants were excluded that did not show an explicit knowledge increase of over 1 standard deviation, creating a group of explicitly instructed participants that actually benefited substantially from this instruction. In doing so, 14 out of the original 32 subjects that received explicit DoC instruction were lost. This means that 44 percent of the subjects was not able to benefit much from the instruction. For SubC, even fewer participants were able to benefit from the explicit instruction: 20 out of 35 subjects had to be removed from the subject sample: a loss of 57 percent of the subjects. There were

no reasons to eliminate subjects from the II conditions: they were not changed, but simply renamed NEK.

One-way ANOVAs were performed to test whether the groups differed significantly at T0, which would introduce a bias in the analyses. No significant differences were found, and the data were considered ready for the intended analyses. Also, despite the loss of subjects for these analyses, the statistical power is still sufficient to detect medium effect sizes.

Interfaces of knowledge

In Table 4.19, the mean scores and standard deviations are presented for both the explicit and implicit measure of grammatical knowledge and broken down according to time of measurement and target structure. Then, two by three univariate analyses of variance were again run for each measure of grammatical development using *Condition* as the between-subjects factor and *progress* as the within-subjects factor. Obviously, as the two contrasted groups were set to differ with respect to explicit knowledge, significant interactions were found for the GJ task for both target structures (for DoC: $F(2, 98) = 20.21$, $p < 0.01$; and for SubC: $F(2, 82) = 24.73$, $p < 0.01$; see also Table 4.20). Quite remarkable, though, was that the explicit knowledge of SubCs for the EK group of learners was not stable, given the deteriorated performance on the GJ task at T2 (see Table 4.19).

TABLE 4.19 *Means and standard deviations for both Doc and SubC GJ task scores and correct use.*

			T0	*T1*	*T2*
Measure	***Condition***	***N***	***M (SD)***	***M (SD)***	***M (SD)***
Degrees of comparison					
GJ task	EK	18	2.06 (1.92)	4.33 (1.53)	4.33 (1.78)
	NEK	35	1.94 (1.78)	2.06 (2.18)	2.49 (2.02)
Correct use	EK	18	6.17 (3.85)	9.06 (5.70)	8.06 (4.53)
	NEK	35	5.09 (4.72)	6.66 (4.54)	7.14 (4.37)
Subordinate clauses					
GJ task	EK	15	1.53 (1.46)	5.07 (1.79)	3.73 (2.89)
	NEK	32	1.44 (1.72)	2.44 (2.26)	3.19 (2.47)
Correct use	EK	15	5.53 (4.22)	7.93 (4.95)	7.80 (4.78)
	NEK	32	6.09 (5.42)	8.19 (5.38)	9.25 (5.94)

TABLE 4.20 *Repeated measures ANOVA for progress and instruction*

Measure and source of variation	*SS*	*df*	*MS*	*F*	*p.*
Degrees of comparison					
GJ task [a]					
progress	54.92	1.75	31.47	26.93	.00
progress x cond.	31.20	1.75	17.88	15.30	.00
Error	104.01	89.00	1.17		
Correct use					
progress	141.57	2	70.79	10.53	.00
progress x cond.	15.74	2	7.87	1.17	.31
Error	685.59	102	6.72		
Subordinate clauses					
GJ task					
progress	124.24	2	62.12	26.39	.00
progress x cond.	37.31	2	18.66	7.93	.00
Error	211.82	90	2.35		
Correct use					
progress	171.80	2	85.90	11.40	.00
progress x cond.	7.88	2	3.94	0.52	.60
Error	678.36	90	7.54		

a – Because the assumption of sphericity was violated, the Greenhouse-Geisser estimate is reported for this measure.

For the Correct Use measure, significant interactions between progress and condition were not observed for either target structure (see Table 4.20). The analyses did turn out significant main effects for all measures. The conclusion must be, though, that even the L2 learners that benefited substantially from the explicit instruction did not demonstrate any advantage in their implicit grammatical performance.

4.3.5 Conclusion

The analyses presented here were intended to test two expectations: (1) explicit instruction is superior only in promoting the development of explicit grammatical knowledge; (2) there will be no differences between the two types of

instruction when it comes to the development of implicit grammatical knowledge. Both hypotheses were confirmed.

Explicit knowledge as measured by the grammaticality judgement task was found to be promoted most by explicit instruction: this finding applied to both target structures. However, there were some differences in explicit development between the two structures. The first difference pertained to the effect of the implicit instruction. According to these data, implicit instruction did not affect explicit DoC development, while it did effectively promote explicit SubC development. Post hoc analyses have demonstrated for the DoC that, at T1 and T2, the EI group outperformed the II group. However, for SubCs, the difference between the EI and II conditions on the GJ task scores was never significant, but both significantly outperformed the control group at T2. A second difference in explicit knowledge gain between the two target structures pertained to the size of the effect. According to Cohen's standards, moderate to large effect sizes were observed between the EI and II groups on the explicit DoC scores. As pointed out, for the SubCs, there were no differences between the two experimental groups, but both outperformed the NI group, and the effect sizes are large. The results suggest that the instruction more effectively promoted explicit knowledge of the SubC than that of the DoC. However, these effect sizes need to be interpreted with some caution, as there was no control group reference for the DoC.

A finding that applied to both explicit and implicit knowledge and to both target structures is that most progress was obtained between T0 to T1. Mostly, the progress obtained between T1 and T2 was small and insignificant, and sometimes slight regression could be observed.

All in all, the hypothesis that explicit knowledge of a particular structure promotes the ability to use that structure in spontaneous situations of language use has to be rejected. No advantage was observed for either structure on the implicit measure. Even when the data where manipulated in order to compare groups that differed substantially in explicit knowledge, no effect of explicit instruction was found. Consequently, this study provides no evidence of an interface between explicit and implicit knowledge.

4.4 FFI and learner characteristics

4.4.1 Introduction

RQ 3: How do developmental readiness, L1 similarity, and Individual Differences affect the success explicit and implicit FFI?

This section is concerned with how learner characteristics affect explicit and implicit progress, and whether they interact with the type of instruction received. A number of individual traits have been considered: developmental readiness, L1 similarity, aptitude, cognitive style, motivation and age. For developmental readiness and L1 background, operationalized as two-level between-subjects factors, 2 by 2 univariate ANOVAs will be used to test how they affect progress individually and in interaction with instruction. Correlation analyses will be used to assess the degree of relationship between the remaining ID variables and explicit and implicit grammatical progress. The overall aim of this section is to find out which variables are related to explicit and implicit progress, possibly in interaction with the kind of instruction received.

Gain scores for the GJ task and correct use in free written response task will be used as indicators of explicit and implicit grammatical progress, respectively. These gain scores are effectively operationalized as the standardized residuals of pairs of observations. Using standardized residuals instead of raw gain scores (or fitted residuals) offers the advantage of having a measure of gain that is standardized. In addition, it is independent of the unit of measurement and differences in scale (which are different for the all the measures used). For explicit knowledge, the gain scores reflect how much progress students obtained from T0 to T1, because explicit knowledge was found to be affected by the instruction. Any progress obtained after the instruction stopped may therefore not be the result of the explicit learning mechanisms intended to monitor. For implicit knowledge, progress from T0 to T2 is investigated. Table 4.21 provides an overview of the different gain scores used as dependent variables in this section.

Subsections 4.3.2 and 4.3.3 focus on developmental readiness and L1 similarity, respectively. In 4.3.4, the relation between knowledge gain and aptitude, cognitive style, motivation and age is investigated.

TABLE 4.21 *The composition of the explicit (EK) and implicit (IK) knowledge gain scores.*

GAIN SCORE	*EXPRESSES:*	*N*
EK DoC gain	Standardized gain score obtained on the grammaticality judgement task for the DoC from T0 to T1.	101
EK SubC gain	Standardized gain score obtained on the grammaticality judgement task for SubCs from T0 to T1.	101
IK DoC gain	Standardized gain score obtained on the free written response task for the DoC from T0 to T2.	76
IK SubC Gain	Standardized gain score obtained on the free written response task for SubCs from T0 to T2.	76
IK cond. SubC Gain	Standardized gain score obtained on the free written response task for conditional SubCs from T0 to T2.	76
IK causal SubC Gain	Standardized gain score obtained on the free written response task for causal SubCs from T0 to T2.	76
Overall EK gain	Composite standardized gain score of the DoC and SubC GJ task gain scores: (EK DoC + EK SubC) / 2 from T0 to T1.	101
Overall IK gain	Composite standardized gain score of the DoC and SubC GJ task gain scores: (IK DoC + IK SubC) / 2 from T0 to T2.	76

4.4.2 Developmental readiness

Preliminary analyses

This section investigates by means of univariate ANOVA whether explicit and implicit gain depend on developmental readiness (DR); and – if possible – whether there is an interaction with the type of instruction received. In 3.5.2, developmental readiness has been operationalized: L2 learners may be considered

TABLE 4.22 *Developmental readiness; number of observations, mean gain scores (standardized, see 4.4.1), and standard deviations for explicit and implicit knowledge, and for both target structures.*

Factor		***EI***		***II (+ NI)*** [a]		***Total***
DR DoC	***N***	***M (SD)***	***N***	***M (SD)***	***N***	***M (SD)***
EK DoC gain						
Ready	33	0.50 (1.05)	34	-0.25 (0.96)	67	0.12 (1.07)
Unready	8	-0.46 (0.37)	6	-0.72 (0.53)	14	-0.57 (0.45)
IK DoC gain						
Ready	27	-0.06 (0.93)	37	0.47 (1.05)	64	0.18 (0.97)
Unready	5	-0.64 (0.61)	7	-0.17 (0.48)	12	-0.96 (0.36)
DR conditional SubC						
EK SubC gain						
Ready	32	0.42 (1.08)	34	0.02 (0.98)	66	0.21 (1.04)
Unready	8	0.04 (1.35)	7	-0.39 (0.62)	15	-0.16 (1.06)
IK cond. SubC gain						
Ready	29	-0.07 (0.72)	36	0.08 (1.09)	65	0.01 (0.94)
Unready	6	0.55 (0.98)	5	-0.83 (1.38)	11	-0.08 (1.33)
DR causal SubC						
EK SubC gain						
Ready	23	0.50 (1.07)	22	0.05 (1.16)	45	0.28 (1.12)
Unready	17	0.12 (1.21)	19	-0.17 (0.59)	36	-0.03 (0.93)
IK causal. SubC gain						
Ready	21	-0.27 (1.08)	24	-0.02 (0.97)	45	0.14 (1.07)
Unready	14	0.27 (1.02)	17	-0.02 (0.84)	31	-0.02 (0.89)

a – For implicit gain scores, the II and NI groups were merged. For explicit gain scores, the NI group was not included.

developmentally ready when they start using the target structure correctly. Naturally, the analyses will be performed separately for the degrees of comparison and subordinate clauses. In addition, given the findings in 4.2, a differentiation will be made for conditional and causal subordinate clauses. Thus, three DR factors have been defined, related to the use of the DoC, conditional SubCs, and causal SubCs: *DR DoC*, *DR cond. SubC*, and *DR causal SubC*.

Making such a division, however, leads to a very uneven distribution of observations over the *unready* and *ready* groups. Table 4.22 provides the number of observations for both explicit gain and implicit gain when a division according

to readiness is made. For explicit gain, the students in the NI condition were not included in the analyses, simply because the number of observations per cell would be too small. Given that instruction has not been found to affect the development of implicit knowledge (see 4.3), the II and NI groups were merged for the analyses pertaining to implicit gain. What stands out is that the number of observations for the unready groups is quite small, especially for the factors *DR DoC* and *DR cond. SubC*. The question that rises is whether there is sufficient statistical power to detect main and interaction effects. In order to obtain sufficient power (.80) to detect large main effect sizes (f = .40 according to Cohen's (1988) standards), a total sample size of 52 is needed, or 26 per group. This means that the total sample size for the investigation of both explicit and implicit gain is adequate, but the total *unready* group sizes for both *DR DoC* and *DR cond. SubC* are too small (see Table 4.22, last column). The same conclusion must be drawn for testing the interaction between developmental readiness and instruction. Detecting large interaction effects (f = .40) at a power of .80 requires a total sample size of 52. In a 2 by 2 design as employed here, this means that each group should contain 13 subjects, a requirement that is frequently not met. Despite the small *unready* group sizes, the intended analyses will be performed. One should keep in mind, though, that if significance fails to occur, this may simply be due to a lack of statistical power.

One question that was addressed before the model proposed here was tested, was whether developmental readiness may be considered an independent predictor of grammatical development. In order to assess this, one-way ANOVAs were conducted for each of the three DR factors with a number of ID and control subject variables. Table 4.23 summarizes the findings. The results show that dividing students into groups according to beginning correct use of the target structure introduced differences between the groups compared. For all three factors, the ready L2 learners outperformed the unready learners on Grammatical Accuracy and the C-test. For Dr DoC, the ready students also did better on Fluency and the Cito ISK test. In addition, L2 learners that were ready for conditional SubCs were significantly younger; and the learners that were ready for causal SubCs did better on the grammatical sensitivity task.

It seems, then, that dividing students into groups according to developmental readiness introduces differences between the groups compared. But perhaps such differences are to be expected, as the construct of developmental readiness expresses 'advancedness' of the L2 learners' linguistic systems. Ready learners were indeed consistently found to outperform unready learners. As these

TABLE 4.23 *One-way ANOVAs for the developmental readiness factors and a number of subject variables.*

	Developmental readiness factor		
	DR DoC	***DR Cond. SubC***	***DR Causal SubC***
Factor	***F (df $_{b,w}$), p.***	***F (df $_{b,w}$), p.***	***F (df $_{b,w}$), p.***
ID variables			-
Memory	-	-	-
Gram. sens.	-	-	5.91 (1,97) *
Age	-	6.01 (1,98) *	-
TJ cogn. style	-	-	-
TJ motivation	-	-	-
Proficiency variables			
Fluency	9.30 (1,99) **	-	-
Gram. Acc.	9.85 (1,99) **	30.93 (1,99) **	15.70 (1,99) **
Spelling	-	-	-
Cito ISK	6.46 (1,99) *	-	-
C-test	12.03 (1,97) **	8.68 (1,97) **	5.58 (1,97) *

* *significant at alpha < .05*

** *significant at alpha < .01*

differences may well be inherent to the construct of DR, they will not be corrected for in the analyses presented here.

Main analyses

Using developmental readiness and instruction as two-level between-subjects factors, univariate 2 by 2 ANOVAs have been run to evaluate how DR affects overall explicit and implicit gain, and how DR interacts with instruction. Table 4.23 shows that, for both explicit and implicit DoC gain, the model predicts a significant amount of between-group variance. For explicit DoC gain, a significant main effect has been found for developmental readiness, which indicates that students that were developmentally ready have obtained more gain than students that were unready. The factor instruction just misses significance in this analysis. The interaction between developmental readiness and instruction is also not

TABLE 4.23 *Degrees of comparison, univariate 2 by 2 ANOVA for instruction and developmental readiness*

Measure and source of variation	***SS***	***df***	***MS***	***F***	***p.***
EK DoC gain					
Corrected model	15.25	3	5.08	5.76	0.00
Intercept	2.41	1	2.42	2.74	0.10
DR DoC	5.87	1	5.87	6.64	0.01
Instruction	2.94	1	2.94	3.33	0.07
DR DoC * Instr.	0.67	1	0.67	0.76	0.38
Error	67.98	77	0.88		
Total	83.24	81			
Corrected total	83.24	80			
IK DoC gain					
Corrected model	13.82	3	4.61	5.51	0.00
Intercept	6.04	1	6.04	7.22	0.00
DR DoC	13.52	1	13.52	16.18	0.00
Instruction	0.00	1	0.00	0.00	0.96
DR DoC * Instr.	0.39	1	0.39	0.47	0.50
Error	60.18	72	0.84		
Total	74	76			
Corrected total	74	75			

significant. For implicit DoC gain, developmentally ready L2 learners also significantly outperform the unready learners. As to be expected, no significance is observed for instruction. The interaction between DR and instruction is also not significant.

The same univariate ANOVAs have been performed for developmental readiness in relation to conditional and causal SubCs, but none of the models was found to predict a significant amount of between-group variance. For explicit SubC gain, irrespective of which DR factor was used, the model was not found to predict any of the variance between the groups: there were no main effects found, and there was also no interaction between developmental readiness and instruction. Because the NI group was not included in these analyses, the lack of significance for the factor instruction was expected (For SubCs, no differences were observed between EI and II; see 4.3.3). The same results were obtained for

both conditional and causal implicit SubC gain: irrespective of the DR factor used, the models did not predict the variance between the groups.

4.4.3 L1 similarity

Preliminary analyses

Another individual characteristic that may interact with instruction is the learner's L1. This section investigates whether learners with similar L1s have an advantage over learners with L1's that are rather different from the L2. L1 similarity has been narrowly defined in specific relation to the two target structures used in this study (see 3.5.3), resulting in two two-level between-subjects factors: *L1 sim DoC* and *L1 sim SubC*. They express whether the target structures are similar or different from in the participants' L1. Univariate ANOVAs will be used to investigate the interaction of these factors with the instruction received for both explicit and implicit gain. As before, the NI group will not be included in these analyses, due to the relatively small number of observations. Table 4.24 describes the input for these analyses. Both L1 similarity factors lead to even distributions across the groups. In order to achieve sufficient statistical power, the same requirements apply as in the previous subsection: 26 subjects per group are needed to detect large main effects, and 13 subjects per group to detect large interaction effects. As there is only one cell frequency below 13 (see Table 4.24), these requirements are sufficiently met.

One-way ANOVAs were performed to investigate whether the groups resulting from a division according to the similarity factors were still sufficiently comparable. Some significant differences were found: for *L1 sim DoC*, the groups differed significantly on the C-test ($F(1,72 = 10.90$; $p < .01$) and on grammatical accuracy ($F(1,74) = 10.84$; $p < .01$); for *L1 sim SubC*, significant differences were observed for aptitude ($F(1,74) = 7.15$; $p < .01$), the Cito ISK test ($F(1,74) = 5.18$; $p < .05$); and spelling ($F(1,74) = 6.96$; $p < .01$). These variables will be introduced as covariates in the following analyses in order to reduce the variance they introduce.

TABLE 4.24 *L1 similarity: number of observations, mean gain scores (standardized, see 4.4.1), and standard deviations for explicit and implicit knowledge, and for both target structures.*

Factor	***EI***		***II***		***Total***	
L1 sim DoC	***N***	***M (SD)***	***N***	***M (SD)***	***N***	***M (SD)***
EK DoC gain						
Similar	17	0.39 (1.02)	15	-0.39 (1.19)	32	0.03 (1.15)
Different	20	0.12 (0.94)	24	-0.28 (0.75)	44	-0.01 (0.85)
IK DoC gain						
Similar	15	0.63 (1.22)	12	-0.16 (0.89)	27	0.28 (1.14)
Different	13	-0.55 (0.78)	22	0.04 (0.96)	35	-0.18 (0.93)
L1 sim SubC						
EK SubC gain						
Similar	21	0.24 (0.98)	24	0.14 (1.13)	45	0.19 (1.04)
Different	16	0.54 (1.37)	15	-0.32 (0.62)	31	0.09 (1.13)
IK SubC gain						
Similar	14	-0.30 (0.93)	21	0.44 (1.06)	35	-0.01 (0.97)
Different	14	0.11 (1.06)	13	-0.11 (0.90)	27	-0.01 (1.00)

TABLE 4.25 *IK DoC gain: univariate 2 by 2 ANOVA for instruction and L1 similarity*

Measure and source of variation	***SS***	***df***	***MS***	***F***	***p.***
Corrected model	20.28	5	4.06	4.93	0.00
Intercept	1.47	1	1.47	1.78	0.19
C-test	4.34	1	4.34	5.27	0.02
Gram. accuracy	1.88	1	1.88	2.28	0.14
L1 sim DoC	0.12	1	0.12	0.14	0.71
Instruction	0.36	1	0.36	0.44	0.51
L1 sim DoC * Instr.	6.43	1	6.43	7.81	0.01
Error	46.10	56	0.82		
Total	66.40	62			
Corrected total	66.38	61			

Main analyses

For both explicit and implicit DoC gain, the overall corrected model was found to explain a significant amount of variation. For explicit DoC gain, a main effect was found for instruction; L1 similarity did not explain a significant amount of variance, nor did the interaction between instruction and L1 similarity. The significant main effect for instruction confirms the findings reported in .3.2. For implicit DoC gain, the results are presented in Table 4.25. The table shows that no significance was found for the individual factors, but the interaction between L1 similarity and instruction was significant: students with L1s that also realize the degrees of comparison by means of morphological suffixation do significantly better on the free written response task when they receive explicit instruction (see Table 4.25). The results did not change when the analysis was run without introducing the covariates.

For both explicit and implicit SubC gain, no significance was observed, irrespective of whether the covariates were introduced, or whether a differentiation was made for conditional and causal SubCs. Nevertheless, these analyses do provide evidence that explicit instruction can be more effective than implicit instruction in promoting implicit knowledge; but only for the DoC, and only if the L1 is similar.

4.4.4 Aptitude, cognitive style, motivation and age

Preliminary analyses

This section focuses on how individual differences affect explicit and implicit progress, and whether there is an interaction with the kind of instruction received. Five relevant ID variables have been identified: two aptitude components – memory and grammatical sensitivity (GS), cognitive style, motivation, and age. In addition, overall aptitude was included, which is a composite score of memory and GS. The degree of relationship between these ID variables and explicit and implicit gain has been explored by means of correlation analyses. Again, gain scores operationalized as standardized residuals have been used as dependent variables. As the effect of the ID variables is evaluated in relation to the kind of instruction received, the NI subjects were excluded for the majority of the analyses presented here: the number of observations is too small to obtain trustworthy correlation coefficients.

Preliminary analyses indicated that there were no significant differences between the groups on any of the ID variables at T0. Also, the degree of interrelationship between the ID variables has been assessed in order to evaluate the extent to which they measure the same constructs. Hardly any significant correlations were found. The two aptitude scores – memory and grammatical sensitivity – were found to be interrelated (R = .41, N = 81, p < .01). The only other correlation found was between grammatical sensitivity and cognitive style (R = -.23, N = 78, p < .05): this latter correlation indicates that students that are considered precise and independent by their teachers, do better on the GS task.

Main analyses

Correlation analyses have been used to investigate the degree of relationship between the ID variables and explicit and implicit gain obtained in the two instruction conditions. In Table 4.26, the Pearson product-moment correlation values are presented for both explicit and implicit DoC gain. For explicit gain, there is a significant negative correlation between gain in the EI condition and age, which means that younger students tend to profit more from the explicit instruction as measured by grammaticality judgements. If gain is measured by means of the free written response task, the EI scores correlate significantly with memory as measured by the paired associates test and overall aptitude. Given that instruction was not found to affect implicit knowledge in previous analyses, a difference between the EI and II conditions is unexpected. However, the correlations between the II scores and memory and aptitude are also rather high (.33 and .32): they just miss out on significance. In addition, due to the small number of observations, the differences between the EI and II correlations would need to be quite large to safely conclude that a variable is differently related to EI and II gain. No significant correlations have been found for grammatical sensitivity, cognitive style and motivation.

The same analyses have been performed for subordinate clauses. The picture that emerges for SubC gain is quite different from that of DoC gain. What stands out in Table 4.27 are the significant correlations of memory, grammatical sensitivity, and overall aptitude with the explicit EI gain scores. In addition, there is again a significant negative correlation between gain in the EI condition and age. However, this time explicit gain in the II condition also correlates negatively with age. Cognitive style shows a significant positive correlation with explicit II gain; and interestingly, there is a negative, although not significant, correlation

TABLE 4.26 *Pearson product-moment correlation coefficients for the degrees of comparison*

	Explicit Doc gain		***Implicit DoC gain***	
	EI	***II***	***EI***	***II***
ID variables	***(N=41)***	***(N=40)***	***(N=32)***	***(N=35)***
Memory	.10	.12	.48**	.33
Gram. sensitivity	-.01	.30	.28	.20
Aptitude	.05	.23	.45*	.32
TJ Cognitive style	.22	-.23	.06	.16
TJ Motivation	.08	-.10	-.15	.16
Age	-.36*	-.18	.01	-.21

* *Significant at p < .05*

** *Significant at p < .01*

TABLE 4.27 *Pearson product-moment correlation coefficients for subordinate clauses*

	Explicit SubC gain		***Implicit SubC gain***	
	EI	***II***	***EI***	***II***
ID variables	***(N=40)***	***(N=41)***	***(N=35)***	***(N=32)***
Memory	.34*	.29	-.08	-.16
Gram. sensitivity	.42*	.18	.19	-.12
Aptitude	.45**	.06	.05	-.16
TJ Cognitive style	-.32	.36*	-.16	-.06
TJ Motivation	-.27	.02	-.27	-.07
Age	-.41**	-.32*	-.25	-.17

* *Significant at p < .05*

** *Significant at p < .01*

with EI gain. The distances between the EI and II correlations are also quite large; at least for GS, Aptitude and Cognitive style, they may fall outside each others confidence intervals. Thus, for these three factors, there may be an interaction with the kind of instruction received. The implicit SubC gain scores do not correlate with any of the ID variables. Further differentiation according to conditional and causal SubCs also did not lead to any significant correlations.

It is difficult to abstract a coherent picture from the correlations presented in tables 4.26 and 4.27. It seems fair to conclude that age is related to explicit gain, irrespective of the kind of instruction provided. Memory and aptitude seem to be

related to DoC gain if measured by means of the free written response task. However, for SubC, the aptitude measures relate to explicit gain in the EI condition.

Individual differences and overall linguistic proficiency

An additional perspective can be obtained by relating the ID variables to a number of other L2 proficiency measures. Five measures are available: the Cito ISK test, the C-test test, Fluency, Grammatical accuracy, and spelling. The Cito ISK test and the C-test test proficiency in controlled contexts (see 3.4.4), while fluency, grammatical accuracy and spelling scores were calculated on the basis of the free written response task data. Table 4.28 shows the correlations. What stands out is that both the Cito ISK test, the C-test, and spelling correlated significantly with all three aptitude measures, while fluency and grammatical accuracy did not (except for a weak correlation between GS and grammatical accuracy). This suggests that aptitude as measured in this study is particularly important in contexts of controlled L2 use. The only other correlation that appeared is the one between cognitive style and spelling: students that were considered structured and precise by their teachers did better on spelling. Motivation and age have not been found to correlate with any of the proficiency measures.

TABLE 4.28 *Pearson product-moment correlation coefficients for subordinate clauses*

	Proficiency measure				
ID variables	***Cito ISK (N=40)***	***C-test (N=41)***	***Fluency (N=35)***	***Gram. ac. (N=32)***	***Spelling (N=35)***
Memory	.28*	.39**	.01	-.13	-.52**
Gram. sensitivity	.41**	.52**	-.07	-.22*	-.25*
Aptitude	.41**	.53**	-.03	-.21	-.47**
TJ Cognitive style	-.12	-.07	.02	-.20	.33**
TJ Motivation	-.07	-.15	-.15	-.02	.13
Age	.16	-.06	-.18	.14	.11

* *Significant at p < .05*

** *Significant at p < .01*

4.4.5 Conclusion

The analyses in this section have focused on how differences between learners affect the development of explicit and implicit grammatical development, and whether this effect is related to the kind of instruction received.

The results presented here lend some support to the hypothesis that developmental readiness is related to effective FFI, in that developmental readiness was found to affect both explicit and implicit progress in the degrees of comparison. Students who already use correct forms of the DoC obtained significantly more progress than students who used no or incorrect forms of the DoC. No interaction could be demonstrated with the kind of instruction students received, either because there is no interaction, or because of the small unready groups causing insufficient statistical power to even detect large effects. For SubC, no evidence was found to suggest that developmental readiness is related to progress. Again, for conditional SubCs, this failure of significance to occur may be due to a lack of statistical power. It should also be noted that developmental readiness was found to be related to a number of proficiency measures, most notably grammatical accuracy and the C-test.

The relation between FFI and L1 background was investigated in a similar way. L1 similarity was found to interact with instruction for implicit DoC gain. Students whose L1 also realizes the degrees of comparison by means of morphological suffixation and who received explicit instruction were found to outperform students that received implicit instruction and students whose L1 is different. This finding is remarkable, because it actually constitutes evidence in favour of an interface between explicit and implicit knowledge. No other significant effects were found for L1 similarity.

In the last subsection, the relation of a number of ID variables with explicit and implicit progress was explored, again in relation to the kind of instruction received. The variables memory, grammatical sensitivity, overall aptitude, cognitive style, motivation and age have been considered. Progress in the DoC was found to be related to memory and overall aptitude, but only if measured implicitly. Explicit progress in the explicit condition was related to age: the younger the students, the more they benefited from the explicit instruction. For SubC, the aptitude measures were related to explicit progress in the EI condition. Cognitive style was related to progress in the II condition, and age was related to progress in both conditions. There were no correlations with implicit progress. In short, age seems to be related to explicit progress of both target structures. For the remainder, progress in the DoC and SubC did not have correlations in

common. All in all, the analyses regarding the ID variables did not provide a coherent picture.

4.5 Summary: differences in the use of DoC and SubC

4.5.1 Introduction

The final research question to be addressed is:

RQ 4: Does the effectiveness of explicit and implicit FFI depend on the complexity of the instructed grammar structure?

In design, this study incorporated two identical, parallel experiments in which two contrasting grammar structures were taught. The motivation for this setup was to put the idea to the test that instruction may have different effects, depending on the complexity of the structure to be taught. Based on previous research, which seemed to suggest a contrast between morphological and syntactic structures, the degrees of comparison and subordination were chosen as targets. As noted before, this contrast was not operationalized as a statistical factor in this research. Rather, all analyses have been conducted and presented separately. This section explores to what extent differences were observed between the two structures with respect to the development of use (4.5.2), the effect of FFI on grammatical development (4.5.3), and the relation between grammatical progress and learner characteristics (4.5.4). In this section, then, an answer to the above mentioned research question will be sought by means of a review of the previously presented findings. As such, this section reads as a summary.

4.5.2 Differences in characteristics of use

In 4.2, a more descriptive analysis of the free written response data was presented, with the specific purpose to explore how second language learners start to make use of the two grammar structures in focus. For both structures, a number of relevant types of realizations were defined, and correct use of the structure was considered in relation to the realizations. For both structures, substantial progress over time was observed in terms of correct use: this observation applied to all contexts of use that were considered. However, if correct use was considered in relation to the different realizations, then there did

seem to be an important contrast. For the DoC, the percentages of correct use were more or less equal across all the different realizations, except when a comparative clause was realized. For the SubCs, though, clear differentiation was observed between the different realizations. The percentages were substantially higher for conditional subordinate clauses as compared to causal subordinate. Similar contrasts were observed for subordination realized with SV (single verb) verb phrases and MV (multiple verbs) verb phrases, and for simple constituent structure clauses and complex constituent structure clauses. Correct use of the SubCs, then, was affected by their realizations, while correct use of the DoC was more or less unaffected by their realizations.

In order to gain insight into how L2 learners move from not using the two target structures to correct use, they were considered in relation to four patterns of use: no use, incorrect use, variable use, and correct use. All participants were classified according to these patterns at each time of measurement. The picture that emerged was very similar for both structures: the participants were found to be quite persistent in how they used the target structures. Over fifty percent of the students did not demonstrate progress in terms of moving from one stage to the next. Also similar for both structures was that a relatively large number of students who demonstrated only correct use at one point, returned to variable use. The conclusion was that for both structures, learning to them in spontaneous situations involves extended periods of incorrect use.

4.5.3 FFI, grammatical development, and structure complexity

In 4.3, two hypotheses were tested by means of univariate repeated measures ANOVA's: explicit instruction (EI) is more effective in promoting explicit knowledge; and explicit and implicit instruction (II) will be equally effective in promoting implicit knowledge. For subordination, the NI condition was also included in the analysis. Comparing the outcomes the analyses, minor differences were observed for the structures.

The analyses were aimed at investigating interactions between instruction (EI, II (and NI)) and progress over time (T0, T1, and T2). They were performed for both measures of progress: the grammaticality judgement task (GJ task) and correct use in the free written response task. For the GJ task, significant overall progress was observed for both structures, as well as an interaction between progress and instruction. However, there were differences between the two structures, which are illustrated by means of Table 4.29. The table shows that at

TABLE 4.29 *Explicit knowledge: post hoc pairwise comparisons at each time of measurement.*

Time of measurement	*Degrees of Comparison*	*Subordination*
T0	EI = II	EI = II = NI
T1	EI > II	EI = II
		EI > NI
		II = NI
T2	EI > II	EI = II > NI

T0, there were no significant differences between the groups. Some differences surface at T1: for the DoC, the EI group was found to have more explicit knowledge than the II group, and this difference persisted at T2. These findings are according to expectation. However, for the SubCs, there are never significant differences between the EI and II group. The progress that was obtained by the EI group was all obtained between T0 and T1, which is why only the EI group differs significant from the NI group at this point. The II group obtained slow but gradual progress, and by T2, they had caught up with the EI group, and both groups significantly outperform the NI group. In sum, implicit instruction led to significant gains in explicit SubC knowledge, but not in explicit DoC knowledge.

The results concerning the interaction between instruction and the development of implicit knowledge are easily summarized: there was none. For both target structures, significant overall progress was observed, and both kinds of instruction were found to be effective. Although the NI group clearly lagged behind the experimental groups, the difference between this group and the experimental conditions was not significant. Even when the EI group was cleared of those students that did not gain explicit knowledge, no interaction was observed. The conclusion was therefore that having explicit knowledge does not constitute an advantage for the acquisition of implicit knowledge.

4.5.4 Individual characteristics and structure complexity

Section 4.4 explored how sources of individual difference may have influenced the effectiveness of instruction. The interaction of explicit and implicit progress with developmental readiness and L1 similarity was explored by means of univariate ANOVAs. For aptitude, attitude and age, correlation analyses were used.

Developmental readiness (DR) was operationalized as emergence of correct use (see 3.5.2). The analyses pertaining to DR revealed clear differences between the two grammar structures, in that DR was found to be related to DoC progress, but not to SubC progress. It was found that students who already used correct forms of the DoC benefit more from the instruction than students who did not yet use correct forms. This effect was observed for both explicit and implicit progress. However, no interaction was found between instruction and developmental readiness, meaning that the beneficial effect of DR applied irrespective of the kind of instruction received. It should be pointed out that the categorization according to whether they had already used correct forms at T0 lead to significant differences between the resulting groups on a number of proficiency measures. Apparently, the construct of DR as it was operationalized in this study expresses to some degree 'advancedness' of the learners interlanguage system. It should also be pointed out that for some of the analyses, insufficient statistical power was achieved, which may have obscured the presence of existing differences.

The analyses pertaining to L1 similarity also revealed a difference between the two structures. L1 similarity was defined in specific relation to the target structures. For the DoC, it divided the first languages of the participating students according to whether they realize comparison morphologically or periphrastically; for subordination, it encoded the default word order of the participants' L1. L1similarity was found to be related only to implicit DoC, and only in interaction with instruction: students with L1's that also express comparison morphologically demonstrated more implicit progress, provided they had had explicit instruction. This result actually constitutes a little bit of evidence in favour of an interface between explicit and implicit knowledge. The effect was only observed for the DoC.

By means of correlation analyses, the relation between explicit and implicit progress in the two target structures was related to aptitude, attitude, and age. For the attitude measures – cognitive style and motivation, and age, no differences were observed with respect to the two grammar structures. The attitude measures were not found to be related to any type of grammatical progress. Age was related to explicit progress of both the DoC and SubCs. There did not seem to be an interaction with age and instruction, although for both the DoC and SubCs the correlation was strongest with EI progress. The analyses pertaining to aptitude did reveal some differences between the two structures. Both memory and grammatical sensitivity (GS) were found to be related to explicit SubC progress, while this relation was absent for explicit DoC progress.

Another finding was that memory was associated to implicit DoC progress, while none of the aptitude measures was implicated in implicit SubC progress. To both observations applied that the associations were strongest for progress in the EI condition. However, the confidence intervals for the correlations presented were quite large, which means that little can be said with certainty about potential interactions with instruction.

4.5.5 Conclusion

In two parallel experiments, the effects of instruction in two contrasting grammar structures were tested. If structure complexity does not affect instruction and acquisition, no differences in outcome should be observed for the two experiments. However, differences occurred frequently. Differentiated use of subordination was observed across different realizations, while such differentiation was largely absent for the DoC. Also, there were differences related to the effectiveness of implicit instruction, developmental readiness, L1 similarity, and aptitude. The question is to what extent these differences are in fact caused by or related to structure complexity. This issue will be taken up in the next chapter, where all findings will be discussed in relation to the theories and earlier research discussed in Chapter 2.